

# **CuGene as a tool to view and explore genomic data**

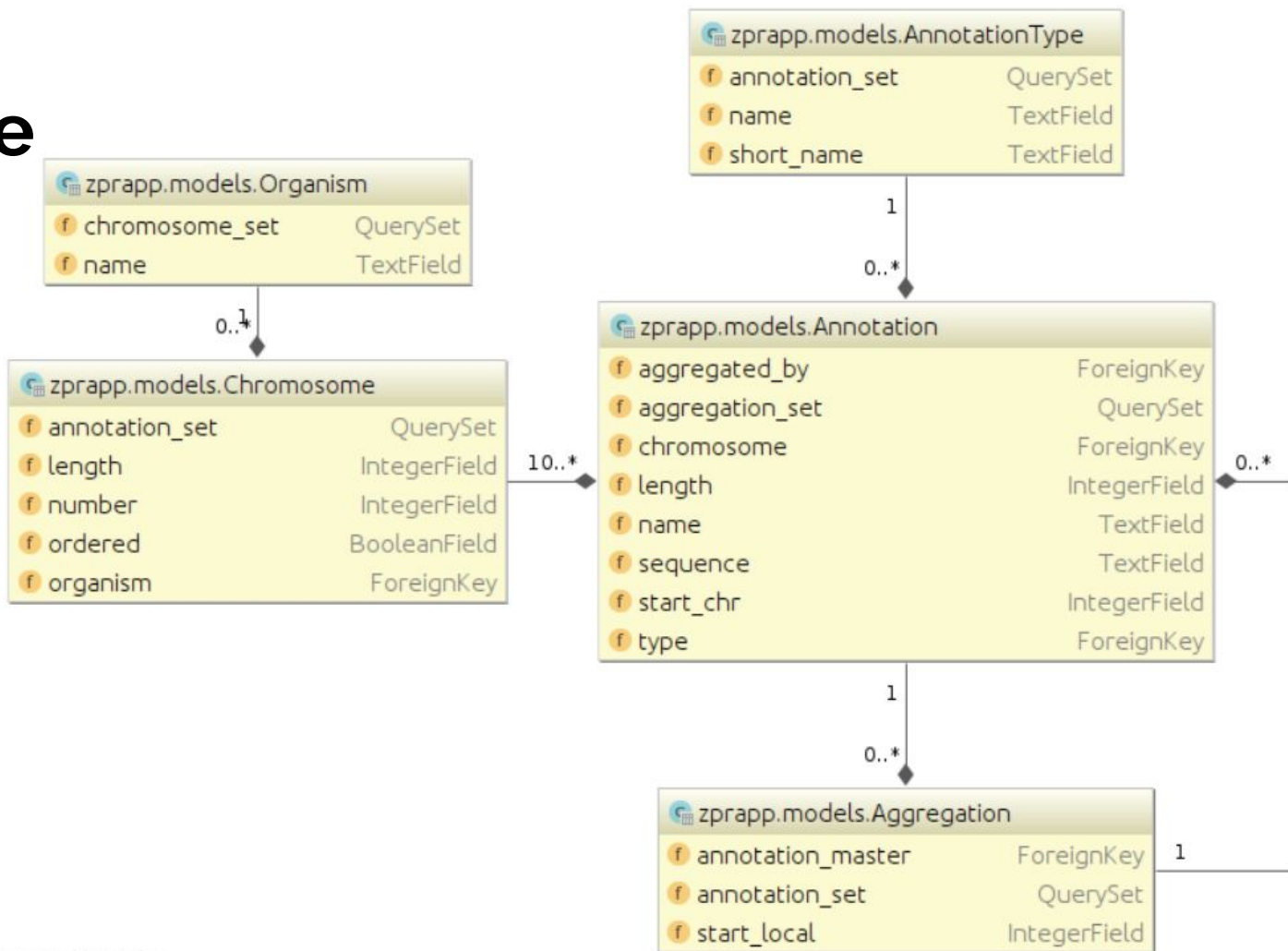
**authors:**

**Michał Haponiuk, Magdalena Pawełkiewicz,  
Zbigniew Przybecki, Robert M. Nowak**

# Software functionality

- display of sequence and annotation tracks
- interactive zoom from megabase to nucleotide resolution
- editing of individual features, supporting manual annotation
- intuitive aligned sequences, elegant to view
- an option for being run as online client-server as well as standalone application
- the possibility to integrate third party applications with CG via HTTP API

# Database



# Input data format

- FASTA
- GFF (general feature format)
- Excel

# Architecture



# Mapped cucumber's genome

Chromosome	chr. length [Mbp]	No. mapped contigs	Mapped length [bp]	Mapped length [%]
1	55	22	32970425	59.9%
2	44	13	23992470	52.2%
3	65	14	39532546	60.8%
4	61	15	24781874	40.6%
5	49	22	26573846	54.2%
6	42	19	29507078	70.2%
7	52	11	18951584	36.4%
$\Sigma$	368	116	196309823	53.3%

- All contigs amount: 8035
- Mapped contigs amount: 116
- Sum of known contig sequence lengths:  
342 288 160 bp
- Sum of mapped contig sequence lengths:  
196 309 823 bp (57%)

Chen, J.-f., Staub, J. E., and Jiang, J., A reevaluation of karyotype in cucumber (*cucumis sativus* L.), Genetic Resources and Crop Evolution 45 (4), 301305 (1998).

# Algorithms

accurate match:

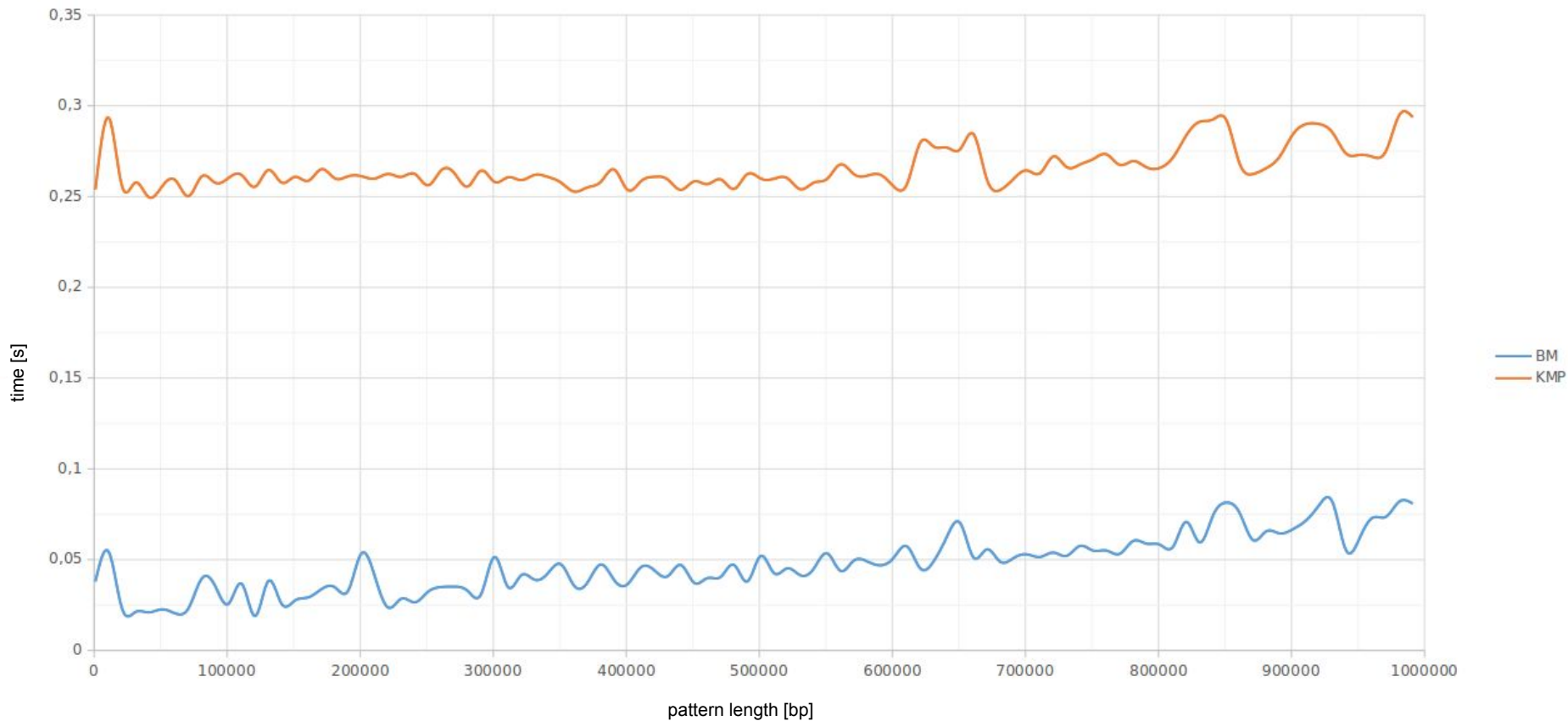
- Knuth-Morris-Prath
- Boyer Moore

searching similarity:

- dynamic programming
  - Smith-Waterman
- heuristic method
  - BLAST

# Algorithms KMP vs BM

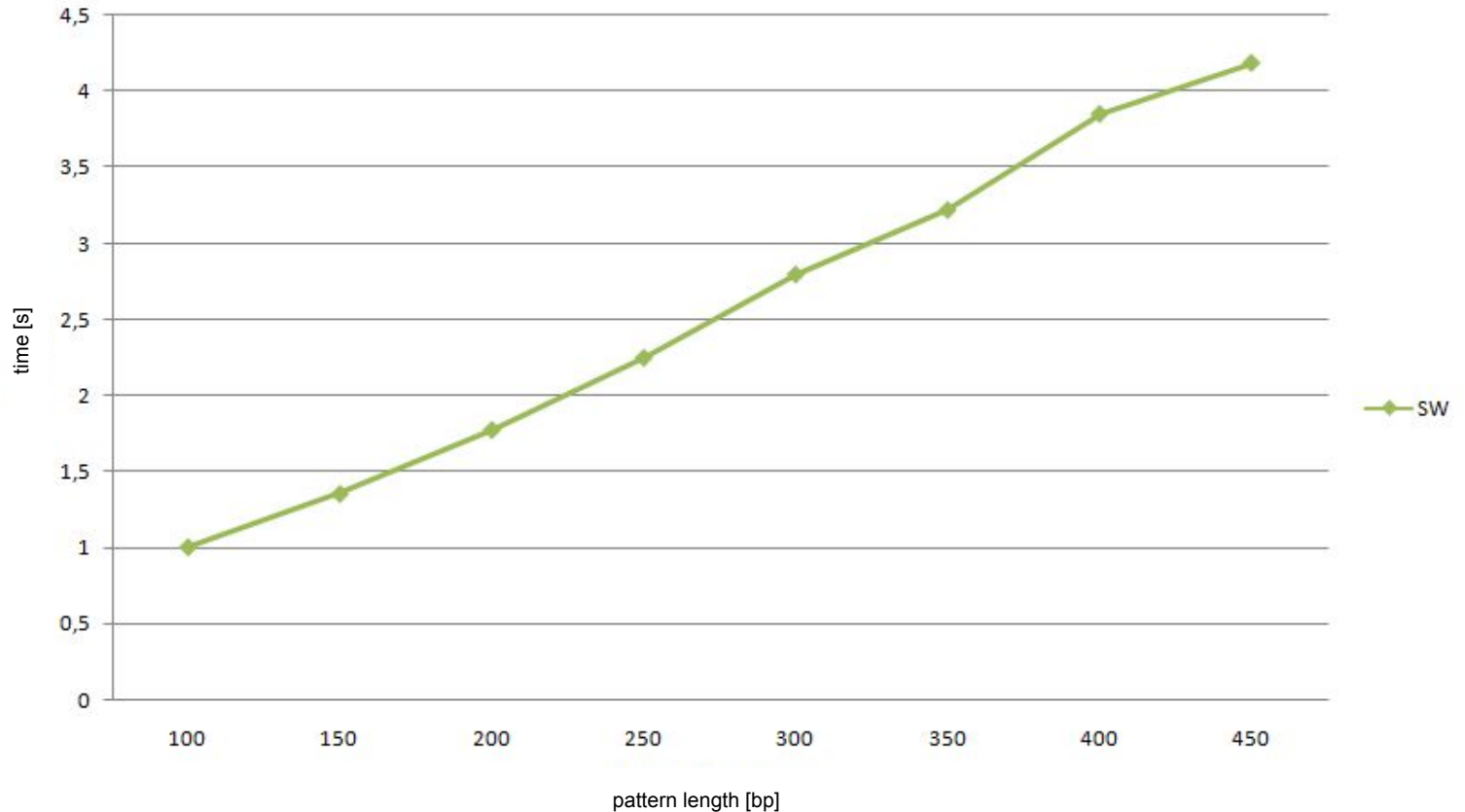
sequence length: 10 000 000 bp





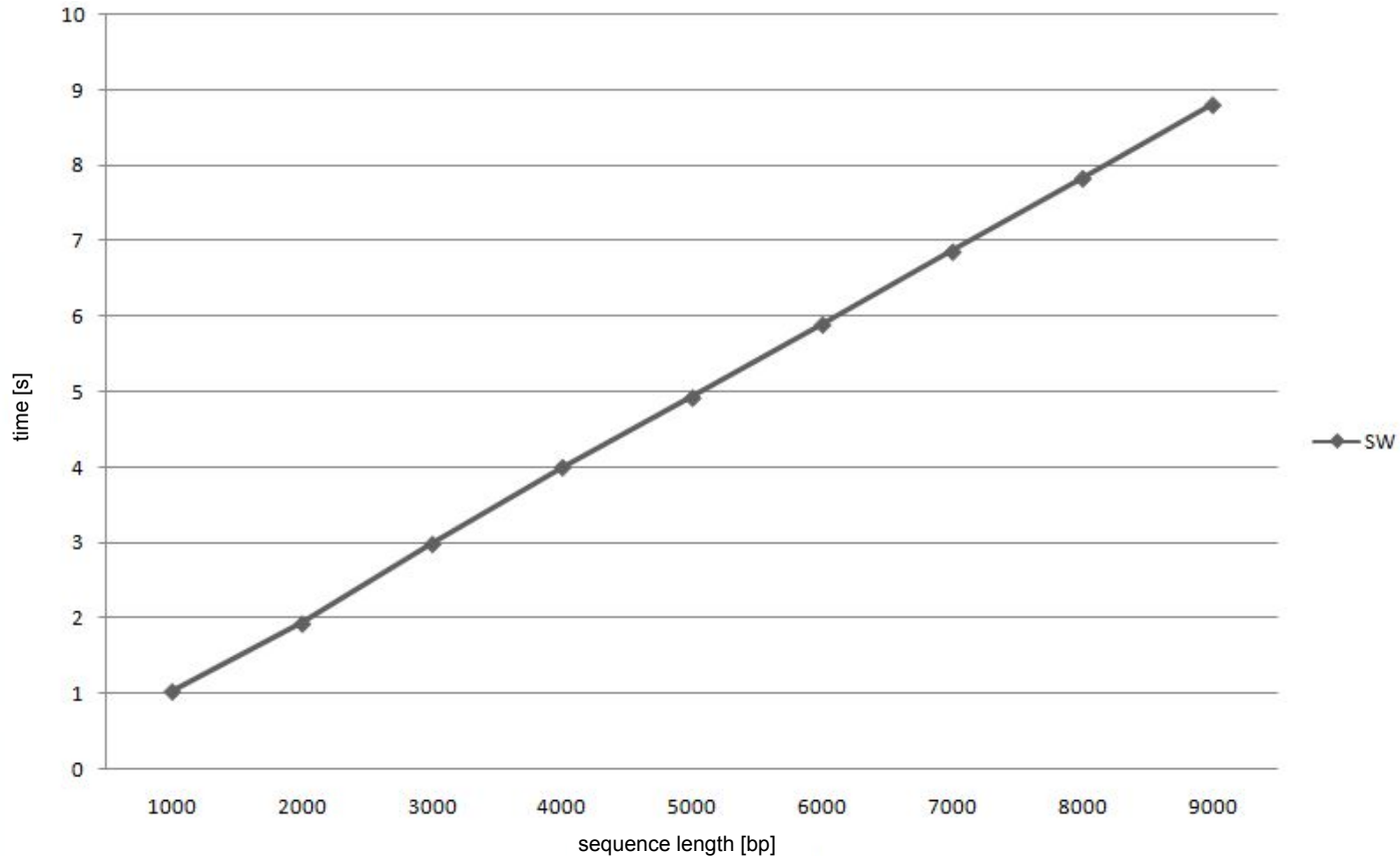
# Smith-Watermann algorithm

sequence length: 5 000 bp



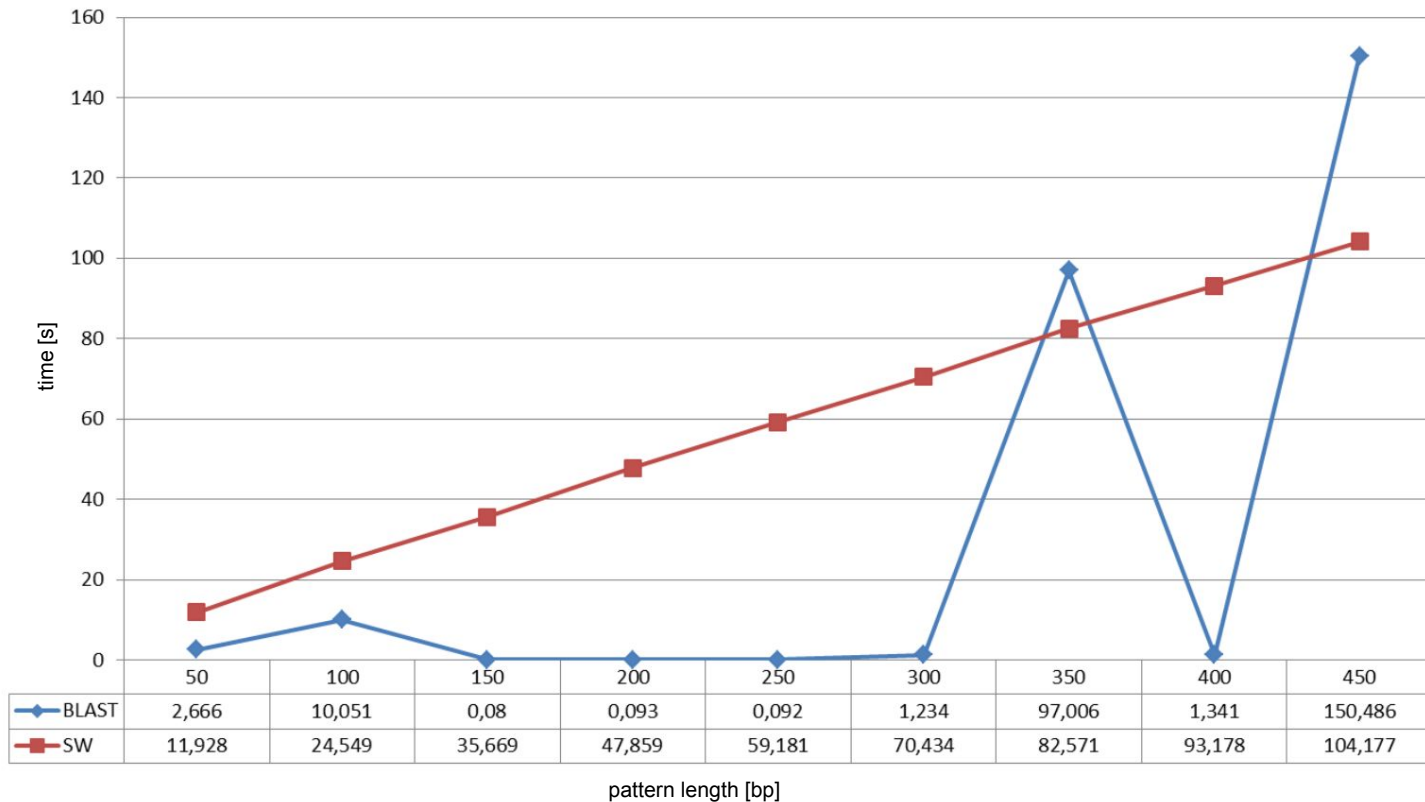
# Smith-Watermann algorithm

pattern length: 500 bp



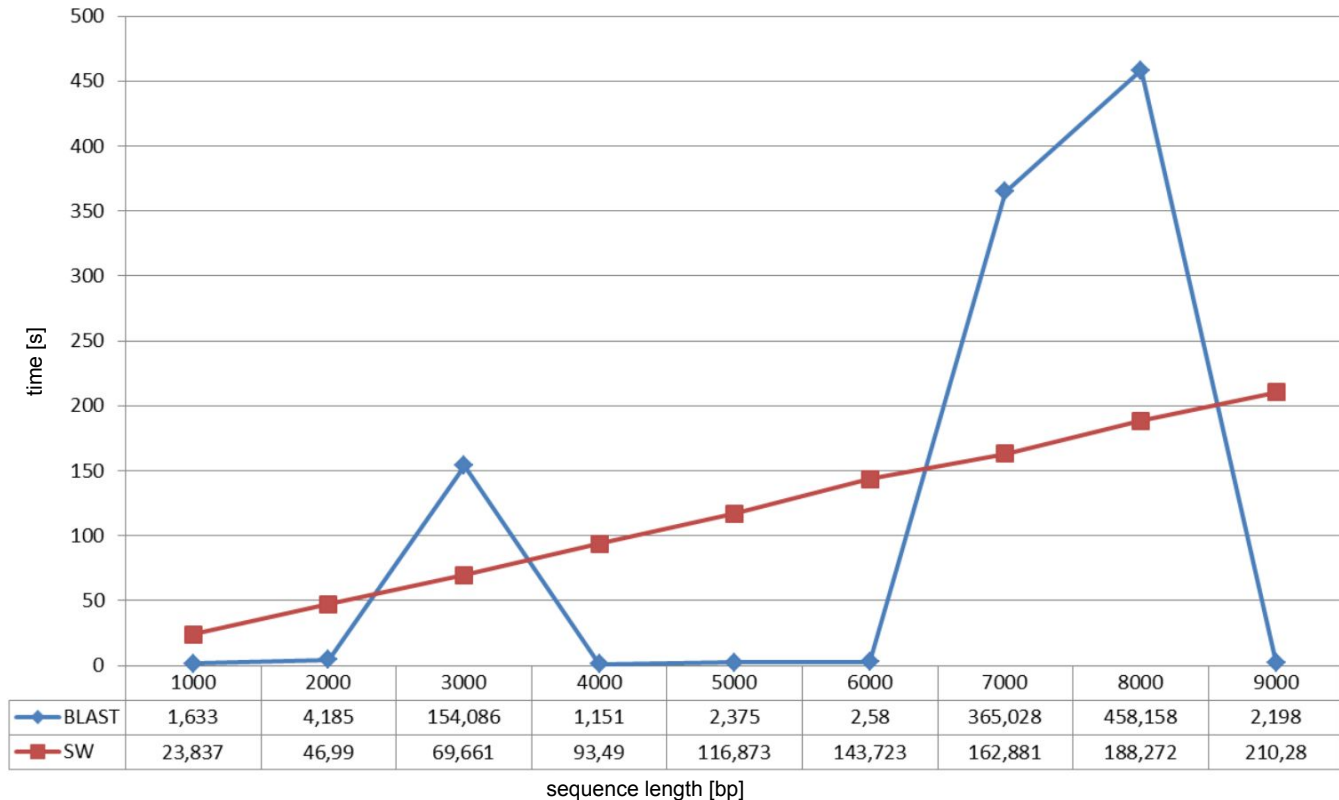
# Smitha-Watermanna vs BLAST

sequence length: 5 000 bp



# Smitha-Watermanna vs BLAST

pattern length: 500 bp



# Genome browsers functionality comparison

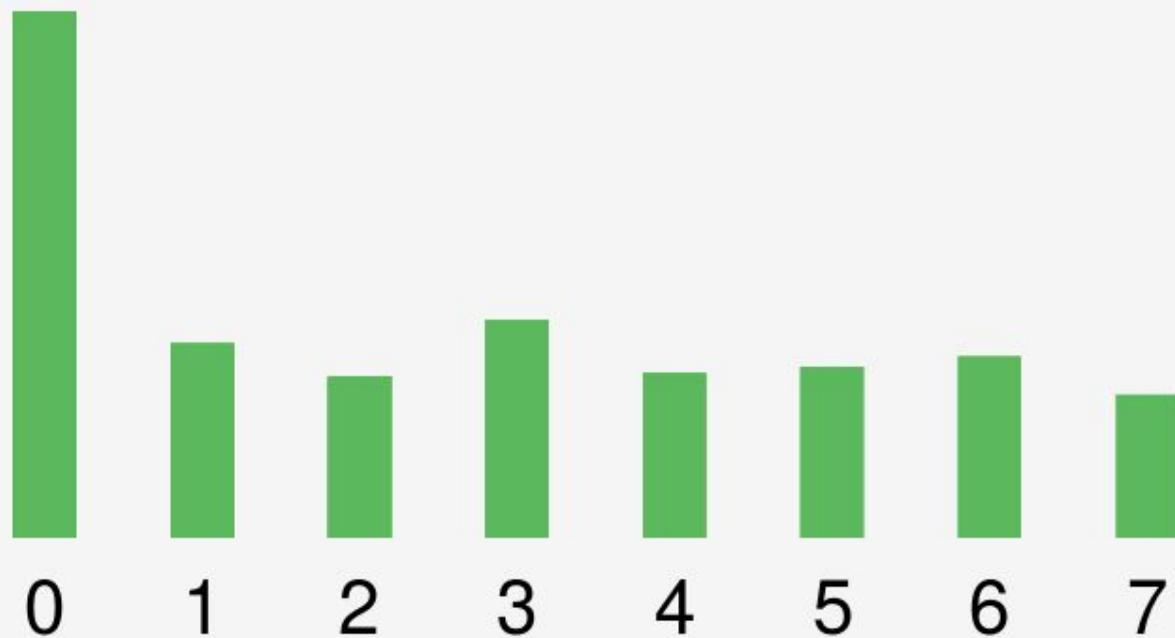
	Embedded data	Custom species		Searching			Frame-work	Stand-alone	Mobile devices
		one	many	BLAST	SW	KMP			
Ensembl	+	+	+	+	-	-	+	+	+
UCSC	+	+	+	+	-	-	+	+	+
Gramene	+	+	+	+	-	-	-	-	+
Gbrowse	-	+	+	+	-	-	+	+	+
Jbrowse	-	+	+	+	-	-	+	+	+
Abrowse	-	+	+	+	-	-	+	+	+
Anno-J	-	+	+	+	-	-	+	+	+
SynBrowse	-	+	+	-	-	-	+	+	-
SynView	-	+	+	-	-	-	+	+	+
LookSeq	-	+	+	+	-	-	+	+	+
Vista	+	-	-	+	-	-	-	-	-
Celera Genome Browser	-	+	-	-	ND	ND	+	+	-
Alamut	+*	-	-	+	ND	ND	-	-	-
GeneWall	+*	-	-	-	-	-	-	+	+
GenomeView	-	+	+	+	-	-	+	+	-
<b>CuGene</b>	-	+	+	+	+	+	+	+	+

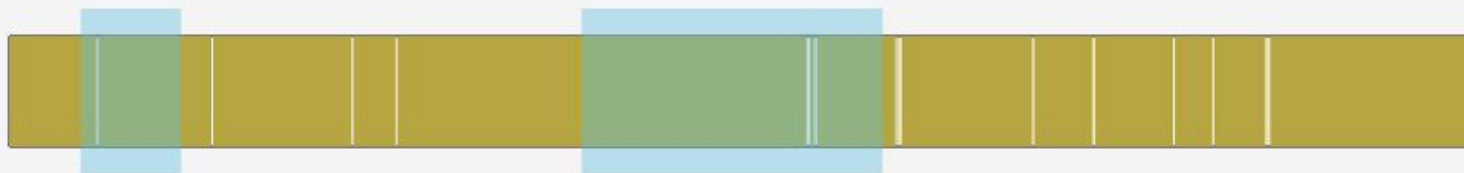


CuGene



0





9.3Mbps 11.1Mbps 12.9Mbps 14.8Mbps 16.6Mbps 18.4Mbps 20.2Mbps 22Mbps 23.9Mbps



50 %

200000

od 9287694 do 23863081

☒ MARKERY

← 14.58Mbps →